Capitolo 5

Le Speranze Condizionate

5.1 La definizione

La definizione di Speranza Condizionata (=SC) si dà usando il teorema di Radon–Nikodym. Ricordo ancora che dati uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ e una sottotribú $\mathcal{G}$ di $\mathcal{F}$ si indica con $\mathbb{P}_{\mathcal{G}}$ la *restrizione* di $\mathbb{P}$ a $\mathcal{G}$, vale a dire la misura di probabilità definita, per ogni insieme B di $\mathcal{G}$, da $\mathbb{P}_{\mathcal{G}}(B) = \mathbb{P}(B)$.

Teorema 5.1.1. *Sia $\mathcal{G}$ una tribú contenuta in $\mathcal{F}$ e sia X una v.a. di $L^1(\mathcal{F})$. Esiste allora una v.a. $Y \in L^1(\mathcal{G})$, unica a meno di equivalenze, tale che per ogni $B \in \mathcal{G}$ risulti*

$$\int_B X \, d\mathbb{P} = \int_B Y \, d\mathbb{P}_{\mathcal{G}} .$$

Dimostrazione. Si scriva la v.a. X come differenza delle sue parti positiva e negativa $X = X^+ - X^-$. Le applicazioni

$$B \mapsto \mathbb{E}(\mathbf{1}_B \, X^+) \qquad \text{and} \qquad B \mapsto \mathbb{E}(\mathbf{1}_B \, X^-)$$

definiscono su $\mathcal{G}$ due misure reali assolutamente continue rispetto a $\mathbb{P}_{\mathcal{G}}$. Esistono allora due funzioni φ^+ e φ^-, uniche a meno di equivalenze, tali che sia, per ogni insieme B di $\mathcal{G}$

$$\mathbb{E}(\mathbf{1}_B \, X^\pm) = \int_B \varphi^\pm \, d\mathbb{P}_{\mathcal{G}} .$$

Basta ora prendere $Y := \varphi^+ - \varphi^-$ per avere l'asserto. □

Definizione 5.1.1. Si dice *Speranza Condizionata* (=SC) della v.a. $X \in L^1(\mathcal{F})$ data la tribú $\mathcal{G} \subseteq \mathcal{F}$ l'unica v.a. di $L^1(\mathcal{G})$, che sarà denotata da $\mathbb{E}(X \mid \mathcal{G})$ o da $\mathbb{E}_{\mathcal{G}}(X)$ tale che, per ogni insieme B di $\mathcal{G}$,

$$\int_B X \, d\mathbb{P} = \int_B \mathbb{E}_{\mathcal{G}}(X) \, d\mathbb{P}_{\mathcal{G}} . \tag{5.1.1}$$

L'unicità si intende nello spazio quoziente $L^1(\mathcal{G})$ e non in $\mathcal{L}^1(\mathcal{G})$, in altre parole, a meno di equivalenze. Se $\mathcal{G}$ è la tribú generata da una v.a. Y su $(\Omega, \mathcal{F}, \mathbb{P})$, $\mathcal{G} = \mathcal{F}(Y)$, si scrive $\mathbb{E}(X \mid Y)$ in luogo di $\mathbb{E}_{\mathcal{F}(Y)}(X)$. ◇

In genere vi saranno molte v.a. $\mathcal{G}$–misurabili che soddisfanno alla (5.1.1); ognuna di esse si dice *versione* della SC. Due qualsiasi versioni della SC sono eguali q.c. (rispetto a $\mathbb{P}$).

Usualmente, e quando ciò non generi confusione, si confonderanno la misura di probabilità $\mathbb{P}$ e la sua restrizione $\mathbb{P}_\mathcal{G}$ alla tribú $\mathcal{G}$ e si scriverà

$$\forall B \in \mathcal{G} \qquad \int_B X \, \mathrm{d}\mathbb{P} = \int_B \mathbb{E}_\mathcal{G}(X) \, \mathrm{d}\mathbb{P} .$$

In qualche caso è possibile estendere la definizione di SC a v.a. che non sono in $L^1(\mathcal{F})$ e che, quindi, non ammettono speranza finita. Per esempio, se X è una v.a. positiva, ma non di L^1, vale a dire che si ha $X \geq 0$ e $\mathbb{E}(X) = +\infty$, non si ha alcuna difficoltà a definire la SC di X.

Il risultato che segue consente di definire le probabilità condizionate data una tribú $\mathcal{G} \subseteq \mathcal{F}$ e di stabilire il legame con le SC.

Teorema 5.1.2. *Esiste un'unica funzione* $\mathbb{P}(\cdot \mid \mathcal{G}) \in L^1(\mathcal{G})$ *detta* probabilità condizionata data la tribú $\mathcal{G}$*, tale che per ogni* $B \in \mathcal{G}$ *e per ogni* $A \in \mathcal{F}$ *valga*

$$\mathbb{P}\left(A \bigcap B\right) = \int_B \mathbb{P}(A \mid \mathcal{G}) \, \mathrm{d}\mathbb{P}_\mathcal{G} ; \tag{5.1.2}$$

vale inoltre

$$\mathbb{P}(A \mid \mathcal{G}) = \mathbb{E}(\mathbf{1}_A \mid \mathcal{G}) . \tag{5.1.3}$$

Dimostrazione. Ponendo, per $A \in \mathcal{F}$ fissato, $\lambda(B) := \mathbb{P}(A \cap B)$ si definisce su $\mathcal{G}$ una misura assolutamente continua rispetto a $\mathbb{P}_\mathcal{G}$. Il teorema di Radon–Nikodym assicura l'esistenza e l'unicità, a meno di equivalenze, di $\mathbb{P}(\cdot \mid \mathcal{G})$. Ponendo $X = \mathbf{1}_A$ nella (5.1.1) si ottiene la (5.1.3) in virtú dell'unicità q.c. sia di $\mathbb{E}(\cdot \mid \mathcal{G})$ sia di $\mathbb{P}(\cdot \mid \mathcal{G})$. $\square$

Anche in questo caso si scriverà, in luogo della (5.1.2),

$$\mathbb{P}\left(A \bigcap B\right)) = \int_B \mathbb{P}(A \mid \mathcal{G}) \, \mathrm{d}\mathbb{P} .$$

5.2 Leggi condizionate

Siamo in grado ora di costruire un modello per la situazione che segue. In uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, sia X una v.a. e sia Y una seconda v.a. la cui legge dipende dal valore x assunto dalla v.a. X. Per esempio, se $x \in [0, 1]$, si può pensare che la v.a. Y rappresenti il numero di teste in n prove bernoulliane indipendenti con probabilità x di successo. Perciò, vorremmo calcolare ciò che intuitivamente potremmo denotare mediante

$$\mathbb{P}(x, B) = \mathbb{P}(Y \in B \mid X = x) ;$$

si noti che la definizione di probabilità condizionata data nei corsi introduttivi di probabilità può non aver senso se l'evento $\{X = x\}$ ha probabilità nulla; nell'esempio appena fatto ciò accade addirittura per ogni $x \in [0, 1]$.

Teorema 5.2.1. *Siano* $(\Omega, \mathcal{F})$ *e* $(\Omega', \mathcal{F}')$ *due spazî misurabili,* $\mathbb{P}$ *una probabilità su* $(\Omega, \mathcal{F})$ *e* $X : \Omega \to \Omega'$ *una funzione misurabile (rispetto alle tribú* $\mathcal{F}$ *e* $\mathcal{F}'$*). Dato un insieme* $B \in \mathcal{F}$ *esiste allora una funzione misurabile* $\varphi_B : \Omega' \to \mathbb{R}$*, tale che per ogni* $A \in \mathcal{F}'$ *sia*

$$\mathbb{P}\left(\{X \in A\} \cap B\right) = \int_A \varphi_B \, \mathrm{d}\mathbb{P}_X \tag{5.2.1}$$

ove $\mathbb{P}_X = \mathbb{P} \circ X^{-1}$ *è la legge immagine di* $\mathbb{P}$ *mediante* X*. Se* ψ *è un'altra funzione che soddisfà alla* (5.2.1) *allora* $\psi = \varphi_B$ *q.c. rispetto a* $\mathbb{P}_X$*. Si pone* $\mathbb{P}(B \mid X = x) := \varphi_B(x)$*.*

Dimostrazione. Ponendo $\mu_B(A) := \mathbb{P}(\{X \in A\} \cap B)$ per ogni $A \in \mathcal{F}'$, si definisce una misura finita su $\mathcal{F}'$ che è assolutamente continua rispetto a $\mathbb{P}_X$. Basta ora applicare il teorema di Radon–Nikodym. □

Si noti che, scritta nella forma,

$$\mathbb{P}(\{X \in A\} \cap B) = \int_A \mathbb{P}(B \mid X = x) \, \mathrm{d}\mathbb{P}_X(x) \,,$$

ove $A \in \mathcal{F}'$ e $B \in \mathcal{F}$, la (5.2.1) è una generalizzazione del teorema delle probabilità totali. In particolare, se $\Omega' = \mathbb{R}$ e se $A = \mathbb{R}$, si ha $\{X \in \mathbb{R}\} = \Omega$ e, dunque,

$$\mathbb{P}(B) = \int_{\mathbb{R}} \mathbb{P}(B \mid X = x) \, \mathrm{d}\mathbb{P}_X(x) = \int_{\mathbb{R}} \mathbb{P}(B \mid X = x) \, \mathrm{d}F(x) \,,$$

ove F è la f.r. della v.a. X.

Esempio 5.2.1. Sia X una v.a. discreta e siano $x_1, x_2, \dots, x_n, \dots$ i valori che essa assume, con probabilità date da $p_j = \mathbb{P}(X = x_j) > 0 \quad (j \in \mathbb{N})$. Mostriamo che

$$\varphi_B(x_j) = \mathbb{P}(B \mid X = x_j) = \frac{\mathbb{P}(B \cap \{X = x_j\})}{\mathbb{P}(X = x_j)} \quad (j \in \mathbb{N}) \,. \tag{5.2.2}$$

Poiché $\mathbb{P}_X$ è concentrata sui singoletti $\{x_j\}$ non occorre specificare φ_B per $x \neq x_j$. Si ponga $\Omega' = \{x_1, x_2, \dots, x_n, \dots\}$, $\mathcal{F}' = \mathcal{P}(\Omega')$. Se A appartiene a $\mathcal{F}'$ e se φ_B è definita dalla (5.2.2), si ha

$$\begin{aligned}
\int_A \varphi_B \, \mathrm{d}\mathbb{P}_X &= \int_{\Omega'} \varphi_B \, \mathbf{1}_A \, \mathrm{d}\mathbb{P}_X = \sum_{n \in \mathbb{N}} \varphi_B(x_n) \, \mathbf{1}_A(x_n) \, \mathbb{P}_X(\{x_n\}) \\
&= \sum_{x_n \in A} \varphi_B(x_n) \, \mathbb{P}(X = x_n) = \sum_{x_n \in A} \mathbb{P}(B \cap \{X = x_n\}) \\
&= \mathbb{P}(B \cap \{X \in A\}) \,.
\end{aligned}$$

Nel caso discreto la definizione del teorema 5.2.1 coincide dunque con quella elementare. ■

Esempio 5.2.2. Siano X e Y due v.a. con legge congiunta assolutamente continua di densità f. In questo caso, per ogni $x \in \mathbb{R}$, l'evento $\{X = x\}$ è trascurabile, sicché non si può definire direttamente la probabilità condizionata

$$\mathbb{P}(Y \in D \mid X = x)\,.$$

Tuttavia, si noti che

$$\begin{aligned}\mathbb{P}(Y \in D \mid x - h < X < x + h) &= \frac{\mathbb{P}(Y \in D, x - h < X < x + h)}{\mathbb{P}(x - h < X < x + h)} \\ &= \left(\int_{x-h}^{x+h} f_1(s)\,\mathrm{d}s\right)^{-1} \left(\int_{x-h}^{x+h} \mathrm{d}s \int_D f(s,t)\,\mathrm{d}t\right),\end{aligned}$$

dove con

$$f_1(s) := \int_{\mathbb{R}} f(s,t)\,\mathrm{d}t$$

si è indicata la densità marginale di X.

Ragionando intuitivamente, si ha che, se f è continua, l'ultima espressione scritta è approssimativamente eguale a

$$\frac{2h}{2hf_1(x)} \int_D f(x,t)\,\mathrm{d}t = \int_D \frac{f(x,t)}{f_1(x)}\,\mathrm{d}t\,.$$

Ciò porta a definire come

$$h(y \mid x) := \frac{f(x,y)}{f_1(x)}$$

la *densità condizionata di Y dato $\{X = x\}$*, o, in breve la densità di Y data X. A rigore $h(y \mid x)$ è definita solo se $f_1(x) \neq 0$; tuttavia, posto

$$S := \{(x,y) \in \mathbb{R}^2 : f_1(x) = 0\}\,,$$

si ha che $\mathbb{P}\left[(X,Y) \in S\right] = 0$. Infatti,

$$\begin{aligned}\mathbb{P}\left[(X,Y) \in S\right] &= \int_S f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_{\{x: f_1(x)=0\}} \mathrm{d}x \int_{\mathbb{R}} f(x,y)\,\mathrm{d}y \\ &= \int_{\{x: f_1(x)=0\}} f_1(x)\,\mathrm{d}x = 0\,.\end{aligned}$$

Si osservi che, se $B \in \mathcal{F}$ e se $X = x$, allora $(x,y) \in B$ se, e solo se, $Y \in B_x$. Questo porta a pensare che sia

$$\varphi_B(x) = \int_{B_x} h(y \mid x)\,\mathrm{d}y\,.$$

Poiché si può scrivere

$$\varphi_B(x) = \int_{\mathbb{R}} h(y \mid x)\,\mathbf{1}_B(x,y)\,\mathrm{d}y\,,$$

si ha intanto che φ_B cosí definita è misurabile. Inoltre, se $A \in \mathcal{B}$,

$$\begin{aligned}
\mathbb{P}(\{X \in A\} \cap B) &= \int\limits_{\{(x,y)\in B, x\in A\}} f(x,y)\, \mathrm{d}x\, \mathrm{d}y \\
&= \int\limits_{\mathbb{R}} \mathbf{1}_A(x) f_1(x)\, \mathrm{d}x \int\limits_{\mathbb{R}} \mathbf{1}_B(x,y)\, h(y \mid x)\, \mathrm{d}y \\
&= \int\limits_{A} f_1(x)\, \mathrm{d}x \int\limits_{B_x} h(y \mid x)\, \mathrm{d}y \\
&= \int\limits_{A} \varphi_B(x)\, f_1(x)\, \mathrm{d}x = \int\limits_{A} \varphi_B(x)\, \mathrm{d}\mathbb{P}_X(x)\,.
\end{aligned}$$

Perciò $\varphi_B(x) = \mathbb{P}(B \mid X = x)$.

L'espressione $f(x,y) = f_1(x)\, h(y \mid x)$ si può leggere nei due versi: da un lato, come si è visto, essa serve a definire la densità condizionata di Y data X, se è nota la densità congiunta f del vettore aleatorio (X,Y), e quindi anche la densità marginale f_1 di X. D'altro canto, se è noto che la v.a. X ha legge di densità f_1, e se si sa che, subordinatamente all'evento $\{X = x\}$, Y ha densità $h(\cdot \mid x)$, allora il vettore aleatorio (X,Y) ha densità $(x,y) \mapsto f(x,y) = f_1(x)\, h(y \mid x)$. Infatti, ponendo, per $B \in \mathcal{B}$,

$$\mathbb{P}(x,B) := \int\limits_{B} h(y \mid x)\, \mathrm{d}y\,,$$

si vede che esiste un'unica misura di probabilità su $\mathcal{B}(\mathbb{R}^2)$ che soddisfaccia, per $A, B \in \mathcal{B}$, a

$$\mathbb{P}(X \in A, Y \in B) = \int\limits_{A} \mathbb{P}(x,B)\, f_1(x)\, \mathrm{d}x = \int\limits_{A} f_1(x)\, \mathrm{d}x \int\limits_{B} h(y \mid x)\, \mathrm{d}y\,.$$

■

Passando alla considerazione delle speranze, si pone il problema di dare una ragionevole definizione della speranza della v.a. Y sapendo che la v.a. X ha preso il valore x; in simboli, si può dare un significato a $\mathbb{E}(Y \mid X = x)$.

Teorema 5.2.2. *Siano $(\Omega, \mathcal{F})$ e $(\Omega', \mathcal{F}')$ due spazî misurabili, $\mathbb{P}$ una probabilità su $(\Omega, \mathcal{F})$, $X : \Omega \to \Omega'$ una v.a. a valori in Ω' e $Y : \Omega \to \mathbb{R}$ una v.a. di $L^1(\mathcal{F})$. Esiste allora una funzione misurabile $\varphi : \Omega' \to \mathbb{R}$ tale che, per ogni $A \in \mathcal{F}'$,*

$$\int\limits_{X^{-1}(A)} Y\, \mathrm{d}\mathbb{P} = \int\limits_{A} \varphi(x)\, \mathrm{d}\mathbb{P}_X(x)\,. \tag{5.2.3}$$

ove $\mathbb{P}_X = \mathbb{P} \circ X^{-1}$ è la legge X. Se ψ è un'altra funzione che soddisfà alla (5.2.3) *allora $\psi = \varphi$ q.c. rispetto a $\mathbb{P}_X$. Si pone $\mathbb{E}(Y \mid X = x) := \varphi(x)$.*

Dimostrazione. Se Y è positiva, si definisce una misura μ su $\mathcal{F}'$, ponendo, per ogni $A \in \mathcal{F}'$,

$$\mu(A) := \int\limits_{X^{-1}(A)} Y\, \mathrm{d}\mathbb{P}\,.$$

L'asserto segue applicando il teorema di Radon–Nikodym alla parte positiva e alla parte negativa di Y. □

La funzione $\mathbb{E}(Y \mid X = x)$ appena introdotta si chiama *speranza di Y condizionata dall'evento* $\{X = x\}$ oppure, piú comunemente, in ispecie nella Statistica Matematica, *funzione di regressione di Y su X*.

Le speranze condizionate includono le probabilità condizionate come caso particolare.

Corollario 5.2.1. *Sia X una v.a. su $(\Omega, \mathcal{F}, \mathbb{P})$ e sia $B \in \mathcal{F}$. Allora vale q.c. rispetto a $\mathbb{P}_X = \mathbb{P} \circ X$,*

$$\mathbb{E}(\mathbf{1}_B \mid X = x) = \mathbb{P}(B \mid X = x)\,.$$

Dimostrazione. Basta porre $Y = \mathbf{1}_B$ nella (5.2.3). □ □

Esempio 5.2.3. Come nell'esempio 5.2.1, siano X una v.a. discreta, $\{x_n : n \in \mathbb{N}\}$ i valori che questaassume, $\mathbb{P}(X = x_n) > 0$ le probabilità con le quali li assume. Si può supporre che $\Omega' = \{x_1, x_2, \dots, x_n, \dots\}$ e $\mathcal{F}' = \mathcal{P}(\Omega')$. Si ponga ora

$$\varphi(x_n) := \frac{1}{\mathbb{P}(X = x_n)} \int_{\{X = x_n\}} Y \, \mathrm{d}\mathbb{P}\,.$$

Allora, per ogni $A \in \mathcal{F}'$, si ha

$$\begin{aligned}\int_{X^{-1}(A)} Y \, \mathrm{d}\mathbb{P} &= \sum_{x_n \in A} \mathbb{P}(X = x_n) \frac{1}{\mathbb{P}(X = x_n)} \int_{\{X = x_n\}} Y \, \mathrm{d}\mathbb{P} \\ &= \sum_{x_n \in A} \mathbb{P}(X = x_n)\varphi(x_n) = \int_A \varphi(x) \, \mathrm{d}\mathbb{P}_X(x)\,,\end{aligned}$$

onde, per ogni $n \in \mathbb{N}$,

$$\mathbb{E}(Y \mid X = x_n) = \frac{1}{\mathbb{P}(X = x_n)} \int_{\{X = x_n\}} Y \, \mathrm{d}\mathbb{P}\,.$$

■

Esempio 5.2.4. Siano X e Y due v.a. con legge congiunta assolutamente continua di densità f e sia $h = h(y|x)$ la densità condizionata di Y data X. Per ogni $A \in \mathcal{B}$, si ha

$$\begin{aligned}\int_{X^{-1}(A)} Y \, \mathrm{d}\mathbb{P} &= \int_{\{(x,y): x \in A\}} y\, f(x,y) \, \mathrm{d}x \, \mathrm{d}y \\ &= \int_A f_1(x) \, \mathrm{d}x \int_{\mathbb{R}} y\, h(y \mid x) \, \mathrm{d}y = \int_A \mathrm{d}\mathbb{P}_X(x) \int_{\mathbb{R}} y\, h(y \mid x) \, \mathrm{d}y\,.\end{aligned}$$

Perciò

$$\mathbb{E}(Y \mid X = x) = \int_{\mathbb{R}} y\, h(y|x) \, \mathrm{d}y\,.$$

È spesso utile la seguente osservazione.

Se è noto che la speranza condizionata è $\varphi(x) := \mathbb{E}(Y \mid X = x)$ e se $\psi := \varphi \circ X$ allora $\mathbb{E}(Y \mid X) = \psi$; infatti dalla definizione di SC, dalla (5.2.3) e dal teorema del cambio di variabile (3.8.2) scende, per ogni $B = X^{-1}(A) \in \mathcal{F}(X)$, con $A \in \mathcal{B}$,

$$\int_{X^{-1}(A)} \mathbb{E}(Y \mid X)\, d\mathbb{P} = \int_B Y\, d\mathbb{P} = \int_A \mathbb{E}(Y \mid X = x)\, d\mathbb{P}_X(x)$$
$$= \int_{X^{-1}(A)} \varphi \circ X\, d\mathbb{P}\,.$$

■

5.3 Proprietà delle speranze condizionate

In questa sezione X denoterà sempre una v.a. di $L^1(\mathcal{F})$ e $\mathcal{G}$ una tribú contenuta in $\mathcal{F}$. Daremo solo le proprietà delle SC data una tribú $\mathcal{G}$, perché le proprietà della speranza della v.a. Y data un'altra v.a. X, $\mathbb{E}(Y \mid X = x)$, o quelle della probabilità condizionata $\mathbb{P}(B \mid X = x)$ sono del tutto analoghe.

Non è inutile ribadire che *tutte le relazioni che saranno date per le SC valgono quasi certamente rispetto alla misura di probabilità* $\mathbb{P}$.

Teorema 5.3.1. *Valgono le seguenti proprietà:*

(a) $\mathbb{E}(\mathbb{E}(X \mid \mathcal{G})) = \mathbb{E}(X)$;

(b) *se* X *è* $\mathcal{G}$*–misurabile* ($X \in L^1(\mathcal{G})$), *allora* $\mathbb{E}(X \mid \mathcal{G}) = X$;

(c) *se* $X = a$ *q.c., allora* $\mathbb{E}(X \mid \mathcal{G}) = a$ *(in particolare* $\mathbb{E}(1 \mid \mathcal{G}) = 1$*)*;

(d) *se* $X \geq 0$, *allora* $\mathbb{E}(X \mid \mathcal{G}) \geq 0$;

(e) *se* $c_1, c_2 \in \mathbb{R}$ *e* $X_1, X_2 \in L^1(\mathcal{F})$, *allora*

$$\mathbb{E}_{\mathcal{G}}(c_1 X_1 + c_2 X_2) = c_1 \mathbb{E}_{\mathcal{G}}(X_1) + c_2 \mathbb{E}_{\mathcal{G}}(X_2)\,,$$

vale a dire che $\mathbb{E}_{\mathcal{G}}$ *è un operatore lineare su* $L^1(\mathcal{F})$;

(f) *se* $X_1 \geq X_2$, *allora* $\mathbb{E}_{\mathcal{G}}(X_1) \geq \mathbb{E}_{\mathcal{G}}(X_2)$;

(g) $|\mathbb{E}_{\mathcal{G}}(X)| \leq \mathbb{E}_{\mathcal{G}}(|X|)$;

(h) *se* $\mathcal{N}$ *è la tribú banale,* $\mathcal{N} := \{\Omega, \emptyset\}$, *allora* $\mathbb{E}_{\mathcal{N}}(X) = \mathbb{E}(X)$.

Dimostrazione. (a) Basta prendere $B = \Omega$ nella (5.1.1). La (b) scende dall'unicità q.c. delle derivate di Radon–Nikodym.

(c) Poiché ogni v.a. che sia q.c. costante è misurabile rispetto alla tribú $\mathcal{N}$, che è inclusa in ogni tribú di sottoinsiemi di Ω, e quindi anche in $\mathcal{G}$, la v.a. $X = a$ q.c. è misurabile rispetto a $\mathcal{G}$; il risultato è dunque contenuto in (a).

(d) Per ogni $B \in \mathcal{G}$, si ha $\int_B X\, d\mathbb{P} \geq 0$.

(e) Per ogni $B \in \mathcal{G}$, si ha

$$\begin{aligned}\int_B \mathbb{E}_{\mathcal{G}}(c_1 X_1 + c_2 X_2)\ \mathrm{d}\mathbb{P} &= \int_B (c_1 X_1 + c_2 X_2)\ \mathrm{d}\mathbb{P} \\ &= c_1 \int_B \mathbb{E}_{\mathcal{G}}(X_1)\ \mathrm{d}\mathbb{P} + c_2 \int_B \mathbb{E}_{\mathcal{G}}(X_2)\ \mathrm{d}\mathbb{P} \\ &= \int_B \{c_1 \mathbb{E}_{\mathcal{G}}(X_1) + c_2 \mathbb{E}_{\mathcal{G}}(X_2)\}\ \mathrm{d}\mathbb{P}\,.\end{aligned}$$

(f) Basta applicare la (d) alla v.a. $X_1 - X_2$ e tenere conto della linearità appena dimostrata.

(g) Basta applicare la (f) e la (e) alla diseguaglianza $-|X| \le X \le |X|$.

(h) $\int_B X\ \mathrm{d}\mathbb{P} = \int_B \mathbb{E}(X)\ \mathrm{d}\mathbb{P}$ se $B = \emptyset$ oppure se $B = \Omega$; l'asserto segue ora dall'unicità q.c. di $\mathbb{E}_{\mathcal{G}}(X)$. □

Osservazione 5.3.1. Le proprietà (g) e (a) dell'ultimo teorema mostrano che l'operatore $X \mapsto \mathbb{E}_{\mathcal{G}}(X)$ è a valori in $L^1(\mathcal{G})$, cioè $\mathbb{E}_{\mathcal{G}} : L^1(\mathcal{F}) \to L^1(\mathcal{G})$. Di piú, esso è lineare, per la proprietà (e), suriettivo, per la proprietà (b) ed è una contrazione, come si vede integrando la (g) ed usando la (a):

$$\|\mathbb{E}_{\mathcal{G}}(X)\|_{L^1(\mathcal{G})} \le \|X\|_{L^1(\mathcal{F})}\,.$$

Ciò rende interessante e naturale lo studio di $\mathbb{E}_{\mathcal{G}}$ come operatore su $L^1(\mathcal{F})$. Tale studio sarà intrapreso piú avanti.

Val la pena di scrivere l'ultima diseguaglianza in una notazione meno precisa, ma, certo, meno pesante,

$$\|\mathbb{E}_{\mathcal{G}}(X)\|_1 \le \|X\|_1\,.$$

Valgono per le SC gli analoghi dei teoremi di Beppo Levi e di convergenza dominata.

Teorema 5.3.2. *Valgono le seguenti affermazioni:*

(a) *Sia (X_n) una successione crescente di v.a. di $L^1(\mathcal{F})$, positive che converge alla v.a. $X \in L^1(\mathcal{F})$, (se, per ogni $n \in \mathbb{N}$, $X_n \ge 0$, $X_{n+1} \ge X_n$, $X_n \xrightarrow[n\to+\infty]{} X$); allora*

$$\lim_{n\to+\infty} \mathbb{E}_{\mathcal{G}}(X_n) = \mathbb{E}_{\mathcal{G}}(X)\,;$$

(b) *se per ogni $n \in \mathbb{N}$, $X_n \ge 0$, allora*

$$\mathbb{E}_{\mathcal{G}}\left(\sum_{n\in\mathbb{N}} X_n\right) = \sum_{n\in\mathbb{N}} \mathbb{E}_{\mathcal{G}}(X_n)\,;$$

(c) *se (B_n) è una successione di insiemi misurabili e disgiunti (per ogni $n \in \mathbb{N}$, $B_n \in \mathcal{F}$, $B_j \cap B_k = \emptyset$, $j \ne k$), allora*

$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} B_n \mid \mathcal{G}\right) = \sum_{n\in\mathbb{N}} \mathbb{P}(B_n \mid \mathcal{G})\,. \tag{5.3.1}$$

Dimostrazione. (a) Per ogni $B \in \mathcal{G}$ e per ogni $n \in \mathbb{N}$, è

$$\int_B X_n \, \mathrm{d}\mathbb{P} = \int_B \mathbb{E}_{\mathcal{G}}(X_n) \, \mathrm{d}\mathbb{P} .$$

Grazie al Teorema 5.3.1(f), la successione $(\mathbb{E}_{\mathcal{G}}(X_n))$ è crescente; essa tende perciò ad una funzione φ necessariamente $\mathcal{G}$–misurabile. Per il teorema di Beppo Levi si ha, perciò, per ogni $B \in \mathcal{G}$,

$$\int_B X \, \mathrm{d}\mathbb{P} = \int_B \varphi \, \mathrm{d}\mathbb{P} ,$$

onde $\varphi = \mathbb{E}_{\mathcal{G}}(X)$.

Le dimostrazioni dei punti (b) e (c) sono semplici applicazioni di (a). □

Osservazione 5.3.2. Si osservi che l'eq. (5.3.1) *non* asserisce che $\mathbb{P}(\cdot \mid \mathcal{G})$ sia una probabilità (si veda la successiva Sezione 5.4). ■

Teorema 5.3.3. *Se (X_n) è una successione di v.a. di $L^1(\mathcal{F})$ che converge q.c. alla v.a. X e se esiste una v.a. $Y \in L^1(\mathcal{F})$ tale che, per ogni $n \in \mathbb{N}$, sia $|X_n| \le Y$, allora*

$$\lim_{n\to+\infty} \mathbb{E}_{\mathcal{G}}(X_n) = \mathbb{E}_{\mathcal{G}}(X) .$$

Dimostrazione. Posto $Y_n := \sup\{|X_j - X| : j \ge n\}$, la successione $\{Y_n\}$ è decrescente e converge a 0. Si osservi che, per ogni $n \in \mathbb{N}$, $\mathbb{E}_{\mathcal{G}}(X_n)$ e $\mathbb{E}_{\mathcal{G}}(X)$ sono in $L^1(\mathcal{G})$. Inoltre

$$|\mathbb{E}_{\mathcal{G}}(X_n) - \mathbb{E}_{\mathcal{G}}(X)| \le \mathbb{E}_{\mathcal{G}}(|X_n - X|) \le \mathbb{E}_{\mathcal{G}}(Y_n) ,$$

sicché basta mostrare che $\mathbb{E}_{\mathcal{G}}(Y_n) \to 0$. Da quanto precede segue che esiste una funzione φ, positiva e $\mathcal{G}$–misurabile, tale che $\mathbb{E}_{\mathcal{G}}(Y_n) \xrightarrow[n\to+\infty]{} \varphi$. Poiché $0 \le Y_n \le 2Y$, il teorema di convergenza dominata dà

$$0 \le \int \varphi \, \mathrm{d}\mathbb{P} \le \int \mathbb{E}_{\mathcal{G}}(Y_n) \, \mathrm{d}\mathbb{P} = \int Y_n \, \mathrm{d}\mathbb{P} \xrightarrow[n\to+\infty]{} 0 ,$$

onde $\varphi = 0$ q.c.. □

Valgono per le SC anche gli analoghi dei lemmi di Fatou.

Teorema 5.3.4. *Sia (X_n) una successione di v.a. di $L^1(\mathcal{F})$, tale che, per ogni $n \in \mathbb{N}$, sia $X_n \ge Y$ con $\mathbb{E}(Y) > -\infty$;*

(a) *se (X_n) è crescente e tende q.c. alla v.a. X, allora*

$$\mathbb{E}_{\mathcal{G}}(X_n) \xrightarrow[n\to+\infty]{} \mathbb{E}_{\mathcal{G}}(X) ;$$

(b) $\liminf_{n\to+\infty} \mathbb{E}_{\mathcal{G}}(X_n) \ge \mathbb{E}_{\mathcal{G}}(\liminf_{n\to+\infty} X_n)$.

Se, invece, (X_n) è una successione di v.a. di $L^1(\mathcal{F})$, tale che per ogni $n \in \mathbb{N}$, sia $X_n \le Y$ con $\mathbb{E}(Y) < +\infty$; allora

(c) *se* (X_n) *è decrescente e tende q.c. alla v.a.* X,

$$\mathbb{E}_{\mathcal{G}}(X_n) \xrightarrow[n\to+\infty]{} \mathbb{E}_{\mathcal{G}}(X)\,;$$

(d) $\limsup_{n\to+\infty} \mathbb{E}_{\mathcal{G}}(X_n) \leq \mathbb{E}_{\mathcal{G}}\left(\limsup_{n\to+\infty} X_n\right)$.

Dimostrazione. Basta dimostrare (a) e (b), perché (c) e (d) scendono da quelle cambiando il segno a tutte le v.a. che intervengono.

(a) Posto $Z_n := X_n - X_1$, la successione (Z_n) è in $L^1(\mathcal{F})$, è positiva e crescente, e ammette come limite la v.a. $X - X_1$. Il teorema di Beppo Levi e la linearità delle SC danno

$$\mathbb{E}_{\mathcal{G}}(X_n) = \mathbb{E}_{\mathcal{G}}(Z_n) + \mathbb{E}_{\mathcal{G}}(X_1) \xrightarrow[n\to+\infty]{} \mathbb{E}_{\mathcal{G}}(X) - \mathbb{E}_{\mathcal{G}}(X_1) + \mathbb{E}_{\mathcal{G}}(X_1) = \mathbb{E}_{\mathcal{G}}(X)\,.$$

(b) Se, per ogni $n \in \mathbb{N}$, si pone $V_n := \inf\{X_j : j \geq n\}$, la successione (V_n) tende crescendo a $V := \liminf_{n\to+\infty} X_n$, onde,

$$\begin{aligned}\mathbb{E}_{\mathcal{G}}(V) &= \mathbb{E}_{\mathcal{G}}(\liminf_{n\to+\infty} X_n) = \lim_{n\to+\infty} \mathbb{E}_{\mathcal{G}}(V_n)\\ &= \liminf_{n\to+\infty} \mathbb{E}_{\mathcal{G}}(V_n) \leq \liminf_{n\to+\infty} \mathbb{E}_{\mathcal{G}}(X_n)\,,\end{aligned}$$

onde l'asserto. □

Le proprietà delle SC che seguono si riferiscono tutte a condizioni di regolarizzazione. Ad esse occorre premettere il seguente lemma di carattere tecnico. Ricordiamo che si dice *atomo* di uno spazio di probabilità $(\Omega, \mathcal{G}, P)$ un insieme $B \in \mathcal{G}$ tale che $\mathbb{P}(B) > 0$ e che se $A \in \mathcal{G}$ e $A \subseteq B$ allora $\mathbb{P}(A) = 0$ oppure $\mathbb{P}(B \setminus A) = 0$.

Lemma 5.3.1. *Sia* B *un atomo di* $(\Omega, \mathcal{G}, \mathbb{P})$ *e sia* X *una v.a. a valori reali su tale spazio. Allora* X *è costante q.c. in* B.

Dimostrazione. Sia B un atomo; Si osservi, innanzi tutto, che si può supporre, che esistano numeri reali x per i quali $\mathbb{P}(B \cap \{X < x\}) = 0$. Infatti, se cosí non fosse, si avrebbe, poiché B è un atomo, $\mathbb{P}(B \cap \{X < x\}) = \mathbb{P}(B)$ per ogni $x \in \mathbb{R}$; in altre parole, sarebbe $X = -\infty$ q.c. in B. Sia, perciò, $x \in \mathbb{R}$ tale che $\mathbb{P}(B \cap \{X < x\}) = 0$; di conseguenza risulta $\mathbb{P}(B \cap \{X < y\}) = 0$ per ogni $y < x$. Posto,

$$k := \sup\{x \in \mathbb{R} : \mathbb{P}(B \cap \{X < x\}) = 0\}\ ,$$

si ha

$$\mathbb{P}(B \cap \{X < k\}) = \mathbb{P}\left(\bigcup_{\substack{q\in\mathbb{Q}\\ r<k}} (B \cap \{X < q\})\right) = 0\,.$$

Se $x > k$, risulta $\mathbb{P}(B \cap \{X < x\}) = \mathbb{P}(B)$, onde $\mathbb{P}(B \cap \{X \geq x\}) = 0$, poiché B è un atomo. Perciò,

$$\mathbb{P}(B \cap \{X > k\}) = \mathbb{P}\left(\bigcup_{\substack{q\in\mathbb{Q}\\ q>k}} (B \cap \{X \geq q\})\right) = 0\,.$$

Pertanto, $X = k$ q.c. su B, cioè $\mathbb{P}(B \cap \{X = k\}) = \mathbb{P}(B)$. □

Siamo ora in grado distabilire l'espressione della SC rispetto ad una tribú $\mathcal{G}$ che sia generata da una partizione misurabile e al piú numerabile di atomi.

Teorema 5.3.5. *Sia B un atomo di $\mathcal{G}$ e sia X una v.a. di $L^1(\mathcal{F})$. Allora*

$$\mathbb{E}_{\mathcal{G}}(X) = \frac{1}{\mathbb{P}(B)} \int\limits_B X \, d\mathbb{P} \qquad \text{q.c. in } B.$$

Dimostrazione. Per il Lemma 5.3.1, $\mathbb{E}_{\mathcal{G}}(X)$ è costante q.c. in B. Perciò, la (5.1.1) dà

$$\mathbb{E}_{\mathcal{G}}(X)\,\mathbb{P}(B) = \int\limits_B X \, d\mathbb{P}\,,$$

cioè l'asserto. □

Corollario 5.3.1. *Se $\pi = \{B_n\}$ è una partizione, finita o misurabile, di Ω in insiemi $\mathcal{F}$–misurabili con $\mathbb{P}(B_n) > 0$, per ogni indice, e se $\mathcal{G}$ è la tribú generata da π, $\mathcal{G} = \mathcal{F}(\pi)$, allora*

$$\mathbb{E}(X \mid \mathcal{G}) := \sum_n \frac{\mathbf{1}_{B_n}}{\mathbb{P}(B_n)} \int\limits_{B_n} X \, d\mathbb{P}\,. \tag{5.3.2}$$

Teorema 5.3.6. *Siano X e Y due v.a. su $(\Omega, \mathcal{F}, \mathbb{P})$ tali che siano integrabili tanto X quanto il prodotto $X\,Y$, $X \in L^1(\mathcal{F})$, $X\,Y \in L^1(\mathcal{F})$; inoltre, Y sia $\mathcal{G}$–misurabile. Allora*

$$\mathbb{E}_{\mathcal{G}}(XY) = Y\,\mathbb{E}_{\mathcal{G}}(X)\,. \tag{5.3.3}$$

Dimostrazione. Si supponga dapprima che Y sia una funzione indicatrice, $Y = \mathbf{1}_B$ con $B \in \mathcal{G}$; in tal caso, per ogni $A \in \mathcal{G}$ si ha

$$\begin{aligned}\int\limits_A \mathbb{E}_{\mathcal{G}}(XY) \, d\mathbb{P} &= \int\limits_A XY \, d\mathbb{P} = \int\limits_{A\cap B} X \, d\mathbb{P} = \int\limits_{A\cap B} \mathbb{E}_{\mathcal{G}}(X) \, d\mathbb{P} \\ &= \int\limits_A \mathbf{1}_B\,\mathbb{E}_{\mathcal{G}}(X) \, d\mathbb{P} = \int\limits_A Y\,\mathbb{E}_{\mathcal{G}}(X) \, d\mathbb{P}\,,\end{aligned}$$

sicché la (5.3.3) vale per le funzioni indicatrici di insiemi di $\mathcal{G}$ e, quindi, per linearità, per le funzioni $\mathcal{G}$–semplici. Ricorrendo al Teorema 5.3.4(a), si dimostra la (5.3.3) per le funzioni $\mathcal{G}$–misurabili positive e, infine, mediante la decomposizione $Y = Y^+ - Y^-$, per tutte le funzioni $\mathcal{G}$–misurabili. □

In particolare se appartengono a $L^1(\mathcal{F})$ tanto X_1 e X_2 quanto il loro prodotto $X_1\,X_2$, vale

$$\mathbb{E}_{\mathcal{G}}\left(X_1\,\mathbb{E}_{\mathcal{G}}(X_2)\right) = \mathbb{E}_{\mathcal{G}}(X_1)\,\mathbb{E}_{\mathcal{G}}(X_2)\,.$$

Se $\mathcal{G}_1$ e $\mathcal{G}_2$ sono due sottotribú distinte di $\mathcal{F}$, in generale gli operatori $\mathbb{E}_{\mathcal{G}_1}$ e $\mathbb{E}_{\mathcal{G}_2}$ non commutano come mostra il seguente

Esempio 5.3.1. Sia $\{B_1, B_2, B_3\}$ una partizione misurabile di Ω e si considerino le tribú

$$\mathcal{G}_1 = \mathcal{F}(\{B_1 \cup B_2, B_3\}), \qquad G_2 = \mathcal{F}(\{B_1, B_2 \cup B_3\})$$

e la v.a. $X = \mathbf{1}_{B_3}$. Allora $\mathbb{E}_{\mathcal{G}_2}\mathbb{E}_{\mathcal{G}_1}X = \mathbb{E}_{\mathcal{G}_2}X$ che non è $\mathcal{G}_1$–misurabile e non può, perciò, coincidere con $\mathbb{E}_{\mathcal{G}_1}\mathbb{E}_{\mathcal{G}_2}X$. ■

Il teorema che segue dà l'esempio di una situazione importante nella quale $\mathbb{E}_{\mathcal{G}_1}$ e $\mathbb{E}_{\mathcal{G}_2}$ commutano.

Teorema 5.3.7. *Siano $\mathcal{G}_1$ e $\mathcal{G}_2$ due sottotribú di $\mathcal{F}$ con $\mathcal{G}_1 \subseteq \mathcal{G}_2$ e sia $X \in L^1(\mathcal{F})$. Allora gli operatori $\mathbb{E}_{\mathcal{G}_1}$ e $\mathbb{E}_{\mathcal{G}_2}$ commutano e vale*

$$\mathbb{E}_{\mathcal{G}_2}\mathbb{E}_{\mathcal{G}_1}X = \mathbb{E}_{\mathcal{G}_1}\mathbb{E}_{\mathcal{G}_2}X = \mathbb{E}_{\mathcal{G}_1}X\,. \tag{5.3.4}$$

Dimostrazione. Poiché $\mathbb{E}_{\mathcal{G}_1}X$ è misurabile rispetto a $\mathcal{G}_1$ e quindi rispetto a $\mathcal{G}_2$, si ha, per la 5.3.1(b)

$$\mathbb{E}_{\mathcal{G}_2}\mathbb{E}_{\mathcal{G}_1}X = \mathbb{E}_{\mathcal{G}_1}X\,.$$

Inoltre, per ogni A in $\mathcal{G}_1$, e dunque anche in $\mathcal{G}_2$, si ha

$$\int_A \mathbb{E}_{\mathcal{G}_1}X \, d\mathbb{P} = \int_A X \, d\mathbb{P} = \int_A \mathbb{E}_{\mathcal{G}_2}X \, d\mathbb{P} = \int_A \mathbb{E}_{\mathcal{G}_1}\mathbb{E}_{\mathcal{G}_2}X \, d\mathbb{P}\,,$$

onde l'asserto. □

Si osservi che se $\mathcal{G}_1 = \mathcal{N}$ e $\mathcal{G}_2 = \mathcal{G}$, la (5.3.3) dà, poiché $\mathbb{E}_{\mathcal{N}} = E$, (Teorema 5.3.1(h)), la 5.3.1(a).

Se $\mathcal{G}_1 = \mathcal{G}_2 = \mathcal{G}$, la (5.3.4) dà $\mathbb{E}_{\mathcal{G}}^2 = \mathbb{E}_{\mathcal{G}}\,\mathbb{E}_{\mathcal{G}} = \mathbb{E}_{\mathcal{G}}$, sicché $\mathbb{E}_{\mathcal{G}}$ è una proiezione da $L^1(\mathcal{F})$ su $L^1(\mathcal{G})$.

Definizione 5.3.1. Una tribú $\mathcal{G} \subseteq \mathcal{F}$ ed una v.a. X su $(\Omega, \mathcal{F}, \mathbb{P})$ si dicono *indipendenti rispetto alla misura di probabilità* $\mathbb{P}$ se sono indipendenti (rispetto a $\mathbb{P}$) $\mathcal{G}$ e la tribú $\mathcal{F}(X) := \{X^{-1}(B) : B \in \mathcal{B}\}$ generata da X. ◇

Teorema 5.3.8. *Se la v.a. X di $L^1(\mathcal{F})$ è indipendente dalla tribú $\mathcal{G}$, si ha*

$$\mathbb{E}_{\mathcal{G}}(X) = \mathbb{E}(X)\,.$$

Dimostrazione. In virtú dell'indipendenza di X e di $\mathcal{G}$, si ha, quale che sia $B \in \mathcal{G}$,

$$\int_B \mathbb{E}_{\mathcal{G}}(X) \, d\mathbb{P} = \mathbb{E}(X\,\mathbf{1}_B) = \mathbb{E}(X)\,\mathbb{E}(\mathbf{1}_B) = \mathbb{E}(X)\,\mathbb{P}(B) = \int_B \mathbb{E}(X) \, d\mathbb{P}\,,$$

onde l'asserto. □

L'ultimo risultato si usa spesso nella seguente forma

Corollario 5.3.2. *Siano X e Y due v.a. indipendenti di $(\Omega, \mathcal{F}, \mathbb{P})$; allora*

$$\mathbb{E}(X \mid Y) = \mathbb{E}(X)\,.$$

Dimostrazione. Infatti, è, per definizione, $\mathbb{E}(X \mid Y) := \mathbb{E}\left(X \mid \mathcal{F}(Y)\right)$ e, inoltre, dire che X e Y sono indipendenti equivale ad affermare che X e la tribú $\mathcal{F}(Y)$ sono indipendenti. $\square$

Il seguente teorema, dovuto a Doob, è spesso utile.

Teorema 5.3.9. *Siano $(\Omega, \mathcal{F})$ e $(\Omega', \mathcal{F}')$ due spazî misurabili e sia $Y : \Omega \to \Omega'$ una v.a.. Per una funzione $\varphi : \Omega \to \mathbb{R}$ sono equivalenti le proprietà:*

(a) *φ è misurabile rispetto a $\mathcal{F}(Y)$ e a $\mathcal{B}$;*

(b) *esiste una funzione misurabile $f : \Omega' \to \mathbb{R}$ tale che $\varphi = f \circ Y$.*

Dimostrazione. Basta dimostrare l'implicazione (a) $\Longrightarrow$ (b). Sia dapprima φ una funzione indicatrice, $\varphi = \mathbf{1}_B$ con $B \in \mathcal{F}(Y)$; allora $B = Y^{-1}(A)$ con $A \in \mathcal{F}'$, sicché, se $f := \mathbf{1}_A$, si ha $f \circ Y = \mathbf{1}_{Y^{-1}(A)} = \mathbf{1}_B = \varphi$. L'asserto è dunque vero se φ è una funzione indicatrice. Se poi φ è una funzione semplice $\varphi = \sum_{j=1}^n \lambda_j \mathbf{1}_{B_j}$ con $\lambda_j \in \mathbb{R}$ e $B_j \in \mathcal{F}(Y)$ $(j = 1, 2, \dots, n)$, allora $\mathbf{1}_{B_j} = f_j \circ Y$ come sopra, con $f_j := \mathbf{1}_{A_j}$ se $B_j = Y^{-1}(A_j)$ e $A_j \in \mathcal{F}'$; quindi, $f = \sum_{j=1}^n \lambda_j f_j$.

Sia ora φ una funzione positiva misurabile rispetto a $\mathcal{F}(Y)$ e a $\mathcal{B}$ e sia (s_n) una successione crescente di funzioni semplici che tende a φ. Per quanto si è appena visto, per ogni $n \in \mathbb{N}$ esiste $f_n : \Omega' \to \mathbb{R}$ misurabile tale che $s_n = f_n \circ Y$. Si definisca $f := \limsup_{n\to+\infty} f_n$. Perciò

$$\varphi = \lim_{n\to+\infty} s_n = \limsup_{n\to+\infty} s_n = \limsup_{n\to+\infty} f_n \circ Y = f \circ Y\,.$$

Il caso generale segue ponendo $\varphi = \varphi^+ - \varphi^-$. $\square$

Teorema 5.3.10. *Sia Y una v.a. su $(\Omega, \mathcal{F}, \mathbb{P})$; e se $\mathcal{G} = \mathcal{F}(Y)$, esiste una funzione boreliana $f_X : \mathbb{R} \to \mathbb{R}$, che dipende da X, tale che sia*

$$\mathbb{E}(X \mid Y) = \mathbb{E}_{\mathcal{F}(Y)}(X) = f_X \circ Y\,.$$

Dimostrazione. È conseguenza immediata del Teorema 5.3.9. $\square$

Vale anche per le SC la diseguaglianza di Jensen.

Teorema 5.3.11. *Sia $\varphi : I \to \mathbb{R}$, con I intervallo (limitato o no, aperto o no) di $\mathbb{R}$, convessa e sia X una v.a. di $L^1(\mathcal{F})$ a valori in I. Se $\mathcal{G}$ è una tribú contenuta in $\mathcal{F}$, vale la diseguaglianza di Jensen*

$$\mathbb{E}_{\mathcal{G}}(\varphi \circ X) \geq \varphi \circ \mathbb{E}_{\mathcal{G}}(X)\,. \tag{5.3.5}$$

Dimostrazione. Sia $I = (a, b)$ con $-\infty \leq a < b \leq +\infty$. In primo luogo risulta $\mathbb{E}_{\mathcal{G}}(X) \in I$ q.c.. Infatti se, per esempio, è $X > a$ q.c., si ha

$$0 \geq \int_{\{\mathbb{E}_{\mathcal{G}} \leq a\}} (\mathbb{E}_{\mathcal{G}}(X) - a)\ \mathrm{d}\mathbb{P} = \int_{\{\mathbb{E}_{\mathcal{G}} \leq a\}} (X - a)\ \mathrm{d}\mathbb{P} \geq 0\,,$$

sicché $\mathbb{P}\left(\mathbb{E}_{\mathcal{G}}(X) \leq a\right) = 0$, cioè $\mathbb{E}_{\mathcal{G}}(X) > a$ q.c.. Com'è noto, il *teorema della linea di supporto*, Teorema 1.17.1, asserisce che esistono due successioni di numeri reali (a_n) e (b_n) tali che, per ogni $x \in I$, si abbia

$$\varphi(x) = \sup\{a_n x + b_n : n \in \mathbb{N}\}\,. \tag{5.3.6}$$

Scende dalla (5.3.6) che, per ogni $n \in \mathbb{N}$, $f \circ X \geq a_n X + b_n$ onde

$$\mathbb{E}_{\mathcal{G}}(\varphi \circ X) \geq a_n \mathbb{E}_{\mathcal{G}}(X) + b_n\,.$$

Di qui, prendendo l'estremo superiore, si ottiene l'asserto. □

Corollario 5.3.3. *Se X è in $L^p(\mathcal{F})$ (con $p \geq 1$) e se $\mathcal{G}$ è una tribú contenuta in $\mathcal{F}$, allora $\mathbb{E}_{\mathcal{G}}(X)$ è in $L^p(\mathcal{G})$ e vale la diseguaglianza*

$$\|\mathbb{E}_{\mathcal{G}}(X)\|_p \leq \|X\|_p \tag{5.3.7}$$

sicché la restrizione di $\mathbb{E}_{\mathcal{G}}$ a $L^p(\mathcal{F})$ è una proiezione e una contrazione su $L^p(\mathcal{G})$.

Dimostrazione. Basta applicare la (5.3.5) alla funzione convessa $\varphi : \mathbb{R} \to \mathbb{R}$ definita da $\varphi(x) := |x|^p$ con $p \geq 1$: $|\mathbb{E}_{\mathcal{G}}(X)|^p \leq \mathbb{E}_{\mathcal{G}}(|X|^p)$. Integrando, e ricordando la 5.3.1(a), si ottiene la (5.3.7). □

5.4 Distribuzioni condizionate regolari

Si è visto che che se (B_n) è una successione di insiemi disgiunti di $\mathcal{F}$ e se $\mathcal{G}$ è una sottotribú di $\mathcal{F}$, allora vale

$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} B_n \mid \mathcal{G}\right) = \sum_{n\in\mathbb{N}} \mathbb{P}\left(B_n \mid \mathcal{G}\right) \qquad \text{q.c.}$$

Ciò non implica che possiamo scegliere $\mathbb{P}(\cdot \mid \mathcal{G})$ in modo che sia una una probabilità per ogni (o quasi ogni) $\omega \in \Omega$. Si supponga, infatti che (B_n) sia una successione di insiemi disgiunti di $\mathcal{F}$; allora

$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} B_n \mid \mathcal{G}\right)(\omega) = \sum_{n\in\mathbb{N}} \mathbb{P}\left(B_n \mid \mathcal{G}\right)(\omega)$$

tranne che per i punti ω di un insieme $N(B_1, B_2, \dots)$ di probabilità nulla. Perciò l'insieme dei punti di Ω nei quali non vale l'additività numerabile è

$$M = \bigcup \{N(B_1, B_2, \dots) : B_1, B_2, \dots \text{insiemi disgiunti di } \mathcal{F}\}\,.$$

In genere tale insieme M è un'unione non numerabile di insiemi di probabilità nulla e non è detto né che appartenga a $\mathcal{F}$, né, se vi appartiene, che abbia probabilità nulla, sicché $\mathbb{P}(\cdot \mid \mathcal{G})$ potrebbe non essere numerabilmente additiva.

Nel seguito della sezione si daranno condizioni che assicurino l'additività numerabile.

Definizione 5.4.1. Siano Y una v.a. su $(\Omega, \mathcal{F}, \mathbb{P})$ e $\mathcal{G}$ una sotto–tribú di $\mathcal{F}$. Si dice che una funzione

$$F : (\Omega, \mathbb{R}) \mapsto [0,1]$$

è *funzione di ripartizione condizionata regolare per* Y *data* $\mathcal{G}$ se sono soddisfatte le due condizioni

(a) $F(\omega, \cdot)$ è una funzione di ripartizione per ogni $\omega \in \Omega$;

(b) per ogni $t \in \mathbb{R}$ è $F(\omega, t) = \mathbb{P}(Y \leq t \mid \mathcal{G})(\omega)$ per quasi ogni $\omega \in \Omega$. $\diamond$

Teorema 5.4.1. *Per ogni v.a.* Y *su* $(\Omega, \mathcal{F}, \mathbb{P})$ *e per ogni sotto–tribú* $\mathcal{G}$ *di* $\mathcal{F}$ *esiste una funzione di ripartizione condizionata regolare per* Y *data* $\mathcal{G}$.

Dimostrazione. Per ogni razionale $q \in \mathbb{Q}$ si scelga una versione della probabilità condizionata $F_q(\omega) := \mathbb{P}(Y \leq q \mid \mathcal{G})(\omega)$. Si scriva l'insieme dei numeti razionali come unione numerabile

$$\mathbb{Q} = \bigcup_{n \in \mathbb{N}} \{q_n\}$$

e si ponga

$$A_{ij} := \left\{\omega \in \Omega : F_{q_j}(\omega) < F_{q_i}(\omega)\right\} \qquad \text{e} \qquad A := \bigcup \{A_{ij} : q_i < q_j\} \, .$$

Poiché la condizione $q_i < q_j$ implica $\mathbb{P}(Y \leq q_i \mid \mathcal{G}) \leq \mathbb{P}(Y \leq q_j \mid \mathcal{G})$ q.c., si ha $\mathbb{P}(A_{ij}) = 0$ e, quindi, $\mathbb{P}(A) = 0$.

Si definisca ora

$$B_i := \left\{\omega \in \Omega : \lim_{n \to +\infty} F_{q_i + (1/n)}(\omega) \neq F_{q_i}(\omega)\right\} \quad \text{e} \quad B := \bigcup_{i \in \mathbb{N}} B_i \, .$$

Ora la successione di v.a.

$$\left(\mathbf{1}_{\{Y \leq q_i + (1/n)\}}\right)_{n \in \mathbb{N}}$$

è decrescente e ha come limite $\mathbf{1}_{\{Y \leq q_i\}}$, sicché

$$F_{q_i + (1/n)} \xrightarrow[n \to +\infty]{} F_{q_i} \qquad \text{q.c.}$$

Pertanto $\mathbb{P}(B_i) = 0$ per ogni $i \in \mathbb{N}$ e $\mathbb{P}(B) = 0$.

Inoltre, posto $C := \{\omega \in \Omega : \lim_{n \to +\infty} F_n(\omega) \neq 1\}$, si ha $\mathbb{P}(C) = 0$ poiché la successione di insiemi $(\{Y \leq n\})$ tende a Ω, sicché $\mathbb{P}(Y \leq n \mid \mathcal{G}) \to 1$ q.c. Similmente, ha probabilità nulla l'insieme $D := \{\omega \in \Omega : \lim_{n \to +\infty} F_n(\omega) \neq 0\}$.

Si definisca $E := A \cup B \cup C \cup D$ e

$$F(\omega, t) := \begin{cases} \lim_{q \to t, q > t} F_q(\omega) \, , & \text{se } \omega \notin E, \\ G(t) \text{ per una qualsiasi f.r. } G, & \text{se } \omega \in E. \end{cases}$$

Si noti che ω non è in A, allora $q \mapsto F_q(\omega)$ è crescente sicché esiste il limite $\lim_{q \to t, q > t} F_q(\omega)$. Per $\omega \notin A \cup B$ e per $t \in \mathbb{Q}$, si ha $\lim_{q \to t, q > t} F_q(\omega) = F_t(\omega)$. Analogamente, per $\omega \notin A \cup C \cup D$ si ha $\lim_{t \to +\infty} F_t(\omega) = 1$ e $\lim_{t \to -\infty} F_t(\omega) = 0$.

F è una f.r. condizionata regolare per Y data $\mathcal{G}$. Si prenda $\omega \notin E$; allora $t \mapsto F(\omega, t)$ è crescente. Se $t < t' < q$ si ha $F(\omega, t) \leq F(\omega, t') \leq F(\omega, q) = F_q(\omega)$; ora $F_q(\omega)$ tende a $F(\omega, t)$ al tendere di q a t, sicché $F(\omega, \cdot)$ è continua a destra.

Se $q \leq t$ si ha

$$F(\omega, t) \geq F(\omega, q) = F_q(\omega) \xrightarrow[q \to +\infty]{} 1$$

sicché $F(\omega, t) \xrightarrow[t \to +\infty]{} 1$. In maniera analoga si mostra che $F(\omega, t) \xrightarrow[t \to -\infty]{} 0$. Pertanto $F(\omega, \cdot)$ è una f.r. e la condizione (a) della Definizione 5.4.1 è soddisfatta.

Per definizione di F si ha, per $q \in \mathbb{Q}$, $\mathbb{P}(Y \leq q \mid \mathcal{G})(\omega) = F(\omega, q) = F_q(\omega)$. Al tendere di q a t decrescendo si ha $F(\omega, t) \to F(\omega, t)$ per ogni $\omega \in \Omega$, a causa della continuità a destra; e, per la convergenza dominata delle SC si ha

$$\mathbb{P}(Y \leq q \mid \mathcal{G}) \xrightarrow[q \to t, q > t]{} \mathbb{P}(Y \leq t \mid \mathcal{G}) \qquad \text{q.c.}$$

Perciò, fissato $t \in \mathbb{R}$ è $\mathbb{P}(Y \leq t \mid \mathcal{G}) = F(\omega, t)$ per quasi ogni $\omega \in \Omega$, ciò che prova la condizione (b) della Definizione 5.4.1 e conclude la dimostrazione. □

Definizione 5.4.2. Siano Y una v.a. sullo spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ e $\mathcal{G}$ una sotto–tribú di $\mathcal{F}$. La funzione $Q : \Omega \times \mathcal{B} \to [0, 1]$ si dice essere una *probabilità condizionata regolare per Y data $\mathcal{G}$* se sono verificate le seguenti condizioni:

(a) $Q(\omega, \cdot)$ è una misura di probabilità per ogni fissato $\omega \in \Omega$;

(b) $Q(\omega, B) = \mathbb{P}(Y \in B \mid \mathcal{G})(\omega)$ q.c. per ogni fissato $B \in \mathcal{B}$. ◇

Teorema 5.4.2. *Per ogni v.a. Y su $(\Omega, \mathcal{F}, \mathbb{P})$ e per ogni sotto–tribú $\mathcal{G}$ di $\mathcal{F}$ esiste una probabilità condizionata regolare per Y data $\mathcal{G}$.*

Dimostrazione. Il Teorema 5.4.1 assicura che esista una f.r. condizionata regolare F per Y data $\mathcal{G}$. Si ponga, per $B \in \mathcal{F}$,

$$Q(\omega, B) = \int_{\{t \in B\}} \mathrm{d}F(\omega, t) \,.$$

Cosí, per ogni $\omega \in \Omega$, $Q(\omega, \cdot)$ è la misura di Stieltjes generata da $F(\omega, \cdot)$, e, come tale, $Q(\omega, \cdot)$ è una misura di probabilità.

Si ponga ora $\mathcal{C} := \{B \in \mathcal{B} : Q(\omega, B) = \mathbb{P}(Y \in B \mid \mathcal{G})(\omega) \text{ q.c.}\}$. Poiché $F(\omega, t) = \mathbb{P}(Y \leq t \mid \mathcal{G})(\omega)$, l'insieme $\mathcal{C}$ contiene tutte le semirette $]-\infty, t]$ con $t \in \mathbb{R}$. Si vede subito che, se $a < b$, allora anche l'intervallo $]a, b]$ appartiene a $\mathcal{C}$; allora vi appartengono anche tutte le unioni finite disgiuinte di intervalli aperti a sinistra e chiusi a destra. Per il teorema della classe monotona si ha $\mathcal{C} = \mathcal{B}$. Dunque, Q è una probabilità condizionata regolare per Y data $\mathcal{G}$. □

5.5 Note al Capitolo 5

Anche le SC furono introdotte da Kolmogorov nella sua monografia del 1933 già piú volte citata. Un approccio differente è presentato da Brown (1976). Lo studio delle SC è stato esteso a ambiti piú generali di quello nel quale si pongono queste lezioni:

o in ispazî mensurali generali, o per v.a. che assumono valori in uno spazio di Banach (Diestel & Uhl, 1977)). Molto studiato è stato il problema della caratterizzazione delle SC; il lavoro pioneristico, in questo campo, è dovuto a Moy (1954). Tra le altre caratterizzazioni (Bahadur, 1955), (Šidák, 1957), (Rota, 1960), (Rao, 1965), (Douglas, 1965), (Olson, 1965), (Ando, 1966), (Pfanzagl, 1967). Nel mio quaderno (Sempi, 1986) ho dato una presentazione unificata, almeno dal punto di vista della notazione, di tali caratterizzazioni.

Si vedano (Pintacuda, 1989) e (Letta & Pratelli, 1997) per ampiamenti sull'importante Teorema 5.3.9.

5.6 Esercizî sul Capitolo 5

5.1. Sullo spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, ove $\Omega =]0,1[$, $\mathcal{F}$ è la tribú di Borel e $\mathbb{P}$ è la misura di Lebesgue, si consideri la partizione

$$\left]0, \frac{1}{4}\right[, \quad \left[\frac{1}{4}, \frac{1}{2}\right[, \quad \left[\frac{1}{2}, \frac{3}{4}\right[, \quad \left[\frac{3}{4}, 1\right[.$$

Si calcoli la SC rispetto alla tribú generata da tale partizione della v.a. $X(t) := \sin 2\pi\, t$.

5.2. (Generalizzazione del Teorema di Bayes) Sia $\mathcal{G}$ una tribú contenuta in $\mathcal{F}$; se G è in $\mathcal{G}$ e A è in $\mathcal{F}$, allora

$$\mathbb{P}(G \mid A) = \frac{\int_G \mathbb{P}(A \mid \mathcal{G})\, \mathrm{d}\mathbb{P}}{\int_\Omega \mathbb{P}(A \mid \mathcal{G})\, \mathrm{d}\mathbb{P}} .$$

Se la tribú $\mathcal{G}$ è generata da una partizione, finita o numerabile, di Ω in insiemi misurabili,

$$\{B_1, B_2, \dots, B_n, \dots\},$$

allora si riottiene la formulazione classica del Teorema di Bayes,

$$\mathbb{P}(B_j \mid A) = \frac{\mathbb{P}(A \mid B_j)\,\mathbb{P}(B_j)}{\sum_n \mathbb{P}(A \mid B_n)\,\mathbb{P}(B_n)} .$$

5.3. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ siano X_1 e X_2 due v.a. indipendenti, entrambe di legge di Poisson di parametro $\lambda > 0$, $X_j \sim \mathcal{P}(\lambda)$ $(j = 1,2)$. Se $Y = X_1 + X_2$, si calcoli $\mathbb{P}(X_1 = k \mid Y)$.

5.4. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ siano X_1 e X_2 due v.a. indipendenti, entrambe di legge esponenziale di parametro $\lambda > 0$, $X_j \sim \Gamma(1, \lambda)$ $(j = 1,2)$. Se $Y = X_1 + X_2$, si calcoli $\mathbb{E}(X_1 \mid Y)$.

5.5. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ siano X_1 e X_2 due v.a. indipendenti entrambe di legge di Bernoulli di parametro p. Posto $Y := X_1 + X_2$ e $\mathcal{G} := \mathcal{F}(Y)$,

(a) si trovino $\mathbb{E}_{\mathcal{G}}(X_1)$ e $\mathbb{E}_{\mathcal{G}}(X_2)$;

(b) sono indipendenti $\mathbb{E}_{\mathcal{G}}(X_1)$ e $\mathbb{E}_{\mathcal{G}}(X_2)$?

5.6. Si dimostrino le diseguaglianze di Markov e di Čebyšev per le SC; se $\alpha > 0$ e se $\mathcal{G}$ è una tribú contenuta in $\mathcal{F}$, allora

$$\mathbb{P}(|X| \geq \alpha \mid \mathcal{G}) \leq \frac{1}{\alpha}\,\mathbb{E}_{\mathcal{G}}(|X|)\,, \quad \text{se } X \in L^1$$
$$\mathbb{P}(|X| \geq \alpha \mid \mathcal{G}) \leq \frac{1}{\alpha^2}\,\mathbb{E}_{\mathcal{G}}(X^2)\,, \quad \text{se } X \in L^2.$$

5.7. Per le SC condizionate vale la diseguaglianza di Schwarz; se $\mathcal{G}$ è una tribú contenuta in $\mathcal{F}$, e se le v.a. X e Y appartengono a L^2, allora

$$\mathbb{E}_{\mathcal{G}}(X\,Y) \leq \mathbb{E}_{\mathcal{G}}^{1/2}(X^2)\,\mathbb{E}_{\mathcal{G}}^{1/2}(Y^2)\,.$$

5.8. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ siano X una v.a. di L^2 e $\mathcal{G}$ una tribú contenuta in $\mathcal{F}$. Posto

$$V_{\mathcal{G}}(X) := \mathbb{E}_{\mathcal{G}}\left[(X - \mathbb{E}_{\mathcal{G}}(X))^2\right]\,,$$

si ha

$$V(X) = \mathbb{E}\,[V_{\mathcal{G}}(X)] + V\left(\mathbb{E}_{\mathcal{G}}(X)\right)\,.$$

5.9. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ sia $\mathcal{G}$ una tribú contenuta in $\mathcal{F}$. Se A è un insieme di $\mathcal{F}$, si definisca un insieme di $\mathcal{G}$ mediante $B := \{\mathbb{E}_{\mathcal{G}}\left(\mathbf{1}_A\right) = 0\}$ e si mostri che, a meno di insiemi trascurabili si ha $B \subseteq A^c$, vale a dire $\mathbb{P}\left(A \cap B\right) = 0$.

5.10. Sia X una v.a. con legge $\Gamma(p, \lambda)$ $(\lambda > 0)$ e S_x una v.a. con legge di Poisson di parametro x, ove x è il valore assunto da X. Si calcoli $\mathbb{P}(S = n)$; è questa una legge nota?

5.11. Sia (X, Y) un vettore aleatorio con legge normale, la cui densità è data da

$$f(x, y) := \frac{1}{2\,\pi\,\sigma_1\,\sigma_2\,\sqrt{1-\rho^2}}\,\exp\left\{-\frac{1}{2\,(1-\rho^2)}\,\left(\frac{x^2}{\sigma_1^2} - \frac{2\,\rho\,x\,y}{\sigma_1\,\sigma_2} + \frac{y^2}{\sigma_2^2}\right)\right\}\,.$$

Si calcolino la densità $h(y \mid x)$ di Y condizionata da $X = x$ e $\mathbb{E}(Y \mid X = x)$.

5.12. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ sia X una v.a. di L^2. Se $\mathcal{G}$ è una tribú contenuta in $\mathcal{F}$, si mostri che

(a) $V\left(\mathbb{E}_{\mathcal{G}}(X)\right) \leq V(X)$;

(b) se $Y := X \wedge \alpha$, allora

$$\mathbb{E}_{\mathcal{G}}\left[\{X - \mathbb{E}_{\mathcal{G}}(X)\}^2\right] \geq \mathbb{E}_{\mathcal{G}}\left[\{Y - \mathbb{E}_{\mathcal{G}}(Y)\}^2\right]\,.$$

5.13. Si calcolino esplicitamente le SC dell'esempio 5.3.1.

5.14. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ siano $X_1, X_2, \dots, X_n$ v.a. indipendenti ed isonome di L^1. Posto, al solito, $S_n := \sum_{j=1}^{n} X_j$, sia $\mathcal{G}_n$ la tribú generata da S_n. Si calcoli la SC $\mathbb{E}_{\mathcal{G}_n}(X_j)$ $(j = 1, 2, \dots, n)$.

5.15. Sia T il triangolo di vertici $(0,0)$, $(1,0)$ e $(1,1)$ e sia (X,Y) il vettore aleatorio avente legge uniforme su T; in altre parole la densità di (X,Y) è data da

$$f(x,y) = 2 \cdot \mathbf{1}_T(x,y)\,.$$

Si calcoli la SC $\mathbb{E}(Y \mid X)$.

5.16. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ siano X e Y v.a. indipendenti di L^1; se Y è centrata, allora

$$\mathbb{E}(|X|) \leq \mathbb{E}(|X+Y|)\,.$$

5.17. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ siano X_1, X_2 e Y v.a. reali tali che i vettori aleatorî (X_1, Y) e (X_2, Y) abbiano la stessa legge.

(a) Se $f : \mathbb{R} \to \mathbb{R}$ è misurabile e positiva (o limitata) si ha, q.c.,

$$\mathbb{E}\left(f \circ X_1 \mid Y\right) = \mathbb{E}\left(f \circ X_2 \mid Y\right)$$

(b) se $g : \mathbb{R} \to \mathbb{R}$ è misurabile e positiva (o limitata), e se

$$\varphi_j\left(X_j\right) := \mathbb{E}\left(g \circ Y \mid X_j\right) \quad (j = 1, 2)\,,$$

allora $\varphi_1 = \varphi_2$ q.c. rispetto alla legge comune di X_1 e X_2.

5.18. X_1 una v.a. con legge uniforme in $(0,1)$. Se $X_1 = x_1$, la v.a. X_2 ha legge uniforme su $(0, x_1)$. In generale, se $X_1 = x_1$, $X_2 = x_2$,..., $X_k = x_k$, allora X_{k+1} ha legge uniforme su $(0, x_k)$. Si trovi $\mathbb{E}(X_n)$.

5.19. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, siano X_1, X_2 e X_3 tre v.a.; sono equivalenti le proprietà:

(a) per ogni $t \in \mathbb{R}$,

$$\mathbb{P}(X_3 \leq t \mid X_1, X_2) = \mathbb{P}(X_3 \leq t \mid X_2) \qquad P\text{–q.c.}\,;$$

(b) per ogni coppia s e t di numeri reali, è

$$\mathbb{P}(X_1 \leq s, X_3 \leq t \mid X_2) = \mathbb{P}(X_1 \leq s \mid X_2)\,\mathbb{P}(X_3 \leq t \mid X_2) \qquad \mathbb{P}\text{–q.c.}.$$

5.20. Nello spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ siano $\mathcal{G}$ una sottotribú di $\mathcal{F}$ e X e Y due v.a. tali che X sia indipendente da $\mathcal{G}$ mentre Y è misurabile rispetto a $\mathcal{G}$. Allora si mostri che

(a) per ogni boreliano $A \in \mathcal{B}^2 = \mathcal{B}(\mathbb{R}^2)$ si ha

$$\mathbb{P}\left((X,Y) \in A \mid \mathcal{G}\right) = \mathbb{P}\left((X,Y) \in A \mid Y\right) \qquad \mathbb{P}\text{–q.c.}\,;$$

(b) per ogni boreliano $D \in \mathcal{B}^2$ si ha

$$\mathbb{P}\left(X+Y \in A \mid \mathcal{G}\right) = \mathbb{P}\left(X+Y \in A \mid Y\right) \qquad \mathbb{P}\text{–q.c.}\,.$$

5.21. Siano X e Y v.a. di L^1. Se $\mathbb{E}(Y \mid X) = X$ e $\mathbb{E}(X|Y) = Y$, allora $X = Y$.

5.22. Siano X e Y due v.a. indipendenti entrambe con legge bernoulliana di parametro p, $X, Y \sim Bi(1,p)$. Introdotta la v.a. Z definita mediante

$$Z := \mathbf{1}_{\{X+Y=0\}},$$

si calcolino le SC $\mathbb{E}(X \mid Z)$ e $\mathbb{E}(Y \mid Z)$. Sono indipendenti?

5.23. Si dimostri che $Y \in L^1(\mathcal{G})$ è la SC $\mathbb{E}_{\mathcal{G}}(X)$ se la relazione

$$\int_A X \, \mathrm{d}\mathbb{P} = \int_A Y \, \mathrm{d}\mathbb{P}$$

vale per ogni insieme A in un π–sistema di sottoinsiemi $\mathcal{C}$ di $\mathcal{F}$ che contiene Ω e che genera la tribú $\mathcal{G}$, $\mathcal{G} = \mathcal{F}(\mathcal{C})$.

5.24. Siano dati uno spazio di probabilità $(\Omega_1, \mathcal{F}_1, P_1)$ ed uno spazio misurabile $(\Omega_2, \mathcal{F}_2)$. Si dice *probabilità di transizione da* $(\Omega_1, \mathcal{F}_1)$ *a* $(\Omega_2, \mathcal{F}_2)$ una funzione $N : \Omega_1 \times \mathcal{F}_2 \to \mathbb{R}_+$ tale che

- per ogni $B \in \mathcal{F}_2$ la funzione $\omega_1 \mapsto N(\omega_1, B)$ sia $\mathcal{F}_1$–misurabile;
- per ogni $\omega_1 \in \Omega_1$ la funzione $B \mapsto N(\omega_1, B)$ è una misura di probabilità.

Un altro nome per N è quello di *nucleo stocastico*. Allora:

(a) la funzione $\omega_1 \mapsto \int_{\Omega_2} f(\omega_1, \omega_2) \, N(\omega_1, \mathrm{d}\omega_2)$ è misurabile rispetto alla tribú $\mathcal{F}_1$;

(b) esiste un'unica misura di probabilità Q su $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ tale che per ogni funzione $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ misurabile rispetto alla tribú prodotto $\mathcal{F}_1 \otimes \mathcal{F}_2$ e limitata, valga

$$\int_{\Omega_1 \times \Omega_2} f \, \mathrm{d}Q = \int_{\Omega_1} \mathrm{d}\mathbb{P}_1(\omega_1) \int_{\Omega_2} f(\omega_1, \omega_2) \, N(\omega_1, \mathrm{d}\omega_2) \, ; \qquad (*)$$

(c) la (*) vale anche quando la funzione misurabile f sia positiva;

(d) una funzione $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ è integrabile rispetto a Q se, e solo se, è

$$\int_{\Omega_1} d\,\mathbb{P}_1(\omega_1) \int_{\Omega_2} |f(\omega_1, \omega_2)| \, N(\omega_1, \mathrm{d}\omega_2) < +\infty \, ;$$

se f appartiene a $L^1(Q)$ allora vale la (*).

Si noti che da questo esercizio si riottiene il teorema di Fubini quando N non dipende da ω_1, vale a dire quando $N(\omega_1, B) = \mathbb{P}_2(B)$ per ogni $\omega_1 \in \Omega_1$ e per ogni $B \in \mathcal{F}_2$, dove $\mathbb{P}_2$ è una misura di probabilità su $(\Omega_2, \mathcal{F}_2)$.

5.25. Nelle condizioni dell'esercizio precedente si considerino le tribú di sottoinsiemi di $\Omega_1 \times \Omega_2$ definite da

$$\widetilde{\mathcal{F}_1} := \{A \times \Omega_2 : A \in \mathcal{F}_1\} \qquad \text{e} \qquad \widetilde{\mathcal{F}_2} := \{\Omega_1 \times B : B \in \mathcal{F}_2\} \, .$$

Si dimostri che per una funzione $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ sono equivalenti le affermazioni

(a) f è misurabile rispetto alla tribú $\widetilde{\mathcal{F}_1}$ (rispettivamente $\widetilde{\mathcal{F}_2}$);

(b) f dipende dalla sola variabile ω_1 (rispettivamente ω_2) e come tale è misurabile rispetto alla tribú $\mathcal{F}_1$ (rispettivamente $\mathcal{F}_1$).